



IBM Applied Data Science Capstone Project

ALTERNATIVE CITIES TO LIVE IN THE US FROM THE TOP 3 STATES OF POPULATION

Description of the data

Abstract

Find the alternative cities to live in the United States of America from the top three states of Population based on the 2019 data.

Srinivasan Sugumar

Contents

Introduction	2
Data	2
I. Sources	2
II. Using the data	2
III. Processing the data	2
References	3

Introduction

The Data is collected from the Wikipedia about the cities of the three states. Now, the data collected is queried for the location and on the popularity of the venue for the proper classification of the cities to be chosen by the personals.

Data

I. Sources

To obtain a list of cities of these states, I referred the Wikipedia for a list of cities in New York, California and Illinois.

I have collected the data from the Wikipedia regarding the population, important venues etc. of the cities of these three states for the year 2019. As I have collected the data from the Wikipedia, only the details available in it are taken into account. Maybe, other results may vary in analysis based on the data sources available to them from other sources. In order to attain the precise analysis and report, we may need more detailed data.

II. Using the data

I have used Mapquest Geocoding (API) to query location data (latitude / longitude) in order to map all the cities from these above specified three states. Also, Foursquare (API) are used to find the list of most popular venues in the cities of these three states.

III. Processing the data

The data of the cities of three states which are available separately in Wikipedia are pooled into a single Pandas Data Frame.

After collecting the data from the Mapquest Geocoding API to query the location of the cities of these three states, using Foursquare API, the list of most popular venues in the cities of these three states are identified.

After the basic information are collected, then the classification K-means clustering algorithm is used for the prediction of the above stated problem.

All the processing details can be found in the Jupiter notebook. All the details of the clustering can be referred in the Jupiter notebook and also in the presentation.

References

- [1.] https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population
- [2.] https://en.wikipedia.org/wiki/List_of_towns_in_New_York
- [3.] https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_California
- [4.] https://en.wikipedia.org/wiki/List_of_municipalities_in_Illinois
- [5.] <https://api.foursquare.com/v2/venues/explore?>
- [6.] <https://www.mapquestapi.com/geocoding/v1/address?>